

Amendments to the Claims

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Previously presented) In a data processing system wherein descriptor vectors associated with a plurality of regions of molecules are stored in a database, a method for generating and storing data characterizing at least one region of said plurality of regions, the method comprising the steps of:

generating an entry comprising

i) an identifier that identifies said at least one region, and

ii) data characterizing a set of axes derived from a property distribution of said at least one region;

iii) a set of axes derived from property distribution information of the at least one region, the set of axes characterizing the at least one region;

generating at least one descriptor vector for the at least one region;

applying a mapping to the descriptor vector associated with said at least one region based on preselected criteria; and

storing said entry in a memory, wherein said key is associated with said entry such that the key indexes the entry for retrieval thereof.

Claims 2-3 cancelled.

4. (Previously presented) The method of claim 1, wherein said property distribution of said at least one region is computed from a convolution with a probe function to a property field.

5. (Previously presented) The method of claim 1, wherein said plurality of descriptor vectors are classified into groups, and wherein said mapping step maps said descriptor

vectors to a space discriminating between said groups of descriptor vectors.

6. (Previously presented) The method of claim 5, wherein said mapping is derived from the steps of:

- generating first data representing differences between said groups of descriptor vectors;
- generating second data representing variations within said groups of descriptor vectors;
- identifying a set of component vectors that maximizes a ratio of variations between groups to the variations within the groups along the component vectors as a discriminant criterion function;

- generating a criterion function for subsets of the component vectors, wherein the criterion function utilizes the first data and the second data;

- for each particular subset of component vectors, calculating a probability value for the criterion functions associated with the particular subset;

- selecting a probability value from probability values for said subsets of component vectors based upon a predetermined criterion;

- identifying the subset of component vectors associated with the selected probability value; and

- generating a mapping to a space corresponding to the subset of component vectors associated with the selected probability value, and storing the mapping for subsequent processing.

7. (Previously presented) The method of claim 6, wherein said first data comprises a matrix ϵ_b representing covariance between said groups of descriptor vectors, and said second data comprises a matrix ϵ_w representing covariance within said groups of descriptor vectors.

8. (Previously presented) The method of claim 7, wherein said criterion function has the general form:

$$f(\hat{w}) = C \left(\frac{\hat{w}^T \varepsilon_b \hat{w}}{\hat{w}^T \varepsilon_w \hat{w}} \right)$$

where \hat{w} is some vector, T indicates a transpose, ε_b is a first data representing covariance, ε_w is a second data representing covariance and C is a constant based upon degrees of freedom in ε_b and ε_w .

9. The method of claim 8, wherein C is determined as follows:

$$C = \frac{1/\text{degrees of freedom in } \varepsilon_b}{1/\text{degrees of freedom in } \varepsilon_w} = \frac{1/(N-1)}{1/(\sum n_i - N)}$$

where N represents the number of groups of descriptor vectors, n_i represents the number of regions, and $\sum n_i$ represents the sum of n_i for the N groups.

10. (Previously presented) The method of claim 7, wherein the step of identifying a set of component vectors that maximizes an F distributed criterion function comprises the substeps of:

determining a set of (eigenvalue, eigenvector) pairs for the matrix ε_w

determining said set of component vectors based upon said set of (eigenvalue, eigenvector) pairs for the matrix ε_w .

11. (Previously presented) The method of claim 10, wherein said statistic for a given subset of component vectors is based upon value of said criterion function for said subset of component vectors.

12. (Previously presented) The method of claim 11, wherein said statistic for a given subset of component vectors has the following form:

$$\Psi_s = C \left(\frac{1}{L_s} \right) \sum f_k$$

where f_k represents the value of the criterion function at a component vector in the given subset, C is a constant, L_s represents the number of f_k values in the given subset of component vectors, and the Σ operation sums over the L_s f_k values in the given subset of component vectors.

13. (Previously presented) The method of claim 12, wherein said a probability value for a particular F-distributed statistic represents a probability value that the particular F-distributed statistic could have been larger by chance.

14. (Previously presented) The method of claim 13, wherein said probability value selected from probability values for said subsets of component vectors is a minimum probability value of said probability values for said subsets of component vectors.

15. (Previously presented) The method of claim 6, wherein said mapping for said at least one descriptor vector performs a loop over each component vector belonging to the subset of component vectors associated with the selected probability; wherein, in each iteration of said loop, dot product of said descriptor vector with a transpose of a unit vector for the given component vector is added to a running sum.

Claims 16-30 cancelled.

31. (Previously presented) The method of claim 1, wherein the at least one descriptor vector is invariant to rotation and translation of the at least one region.

32. (Previously presented) The method of claim 31, wherein the set of axes is derived from principal axes of second moments of a region of the property distribution information.

33. (Previously presented) The method of claim 6, wherein the probability value is obtained by treating the ratio as an F-distributed statistic.

Claims 34 - 35 cancelled.